W 1

# Nonparametric Tests for Changes in the Cyclical Sensitivity of Prices

by J. Armstrong



The Research Department of the Bank of Canada undertook the preparation of the study "Price Flexibility and Business Cycle Fluctuations in Canada - A Survey" for the Royal Commission on the Economic Union and Development Prospects for Canada. This paper was a working document for the study. The views expressed are those of the author(s): no responsibility for them should be attributed to the Bank of Canada.

The author would like to thank colleagues at the Bank for their contributions and comments. However, the author takes full responsibility for the content of the paper.

**NONPARAMETRIC TESTS FOR CHANGES IN THE CYCLICAL SENSITIVITY OF PRICES**

In a recent note (Armstrong 1983) I commented on the methodology used by Cagan (1975) to test for intertemporal changes in the cyclical sensitivity of prices. In that note I suggested that the Cagan's essentially descriptive methodology could be more effectively applied in the Canadian context if combined with some formal statistical analysis. In particular I suggested that some nonparametric statistical tests could be used to evaluate hypotheses concerning changes in various aspects of price behaviour over time. The purpose of this note is to provide some details concerning the computation of those test statistics and their particular relevance.

Cagan used data for over one thousand U.S. product prices. For each of a number of post-World War II recession periods, using all the series available, he set up cumulative density functions for a measure of average price change normalized on inflation. He then computed various statistics (mean, variance, etc.) describing those empirical density functions and commented on the way in which the characteristics of the distributions changed over time. He did not comment on the statistical significance of any of the changes he observed.

Behind the Cagan analysis is the implicit assumption that the product prices he considered represent the entire population of relevant prices in the economy. In the Canadian context far fewer than one thousand price series are available over a time span of useful length. In this context the question of whether or not changes in the cyclical behaviour of observed prices would hold up if data were available for all product prices in the economy is a much more compelling one. Discussion of differences between Cagan-style cumulative density functions should be supplemented with results of statistical tests of the significance of observed differences.

A general caveat is necessary before I discuss particular hypotheses concerning differences between cumulative density functions of price changes, test statistics appropriate to the hypotheses and their economic interpretation. In order to use any statistical tests one must assume that the set of product prices used represents a sample drawn randomly from a population of such prices. A sample of prices selected based on the availability of data is patently non-random. From a practical point of view this problem is of secondary importance. More significant is the issue of whether or not the set of prices available includes disproportionate representation of products whose prices, according to a

*This paper is one of the series of working papers for "Price Flexibility and Business Cycle Fluctuations in Canada - A Survey", a study prepared by the Research Department of the Bank of Canada for the Royal Commission on the Economic Union and Development Prospects for Canada. These research papers were all completed in early 1984.*

priori evidence, are, for example, more sensitive to cycles than most prices in the economy. In the Canadian context, samples of sub-aggregate Consumer Price Index and Industry Selling Price Index series have been selected by including only those series for which data are available as early as the late 1940s and early 1950s. These sets of series are spread reasonably uniformly across aggregate groups, and do not include disproportionate representation of any particular product group. Thus, the assumption that each set of series represents a random sample from a population, while wrong, is probably not wrong in any important way.

Now consider, for example, a set of Consumer Price Index sub-aggregate series selected on the basis of data availability. Suppose we make the plausible assumption that this set of series represents a random sample of Consumer Price Index series. Following the Cagan methodology we compute an appropriately normalized measure of price change for each series over each of the k recessions for which data are available. For each recession a cumulative density function of the price-change measures can be set up. What statistical tests can be used to examine the significance of differences between these sample densities?

Two sorts of questions may be of practical interest. Each cumulative density function constructed is an empirical density based on a sample drawn from a population of product-price changes. The cumulative density function for each population is, of course, unknown. First, does the sample of prices examined provide evidence against the hypothesis that groups of two or more of the population densities are identical? If one denotes by $F_i(x)$ the cumulative density function for the population in recession i, we are interested in testing hypotheses of the general form

$$H_0: F_{i_1}(x) = F_{i_2}(x) = \ldots = F_{i_n}(x), \text{ for all } x, \qquad (1)$$

where $i_1, i_2 \ldots i_n$ are unique and $n \leq k$. The second type of question of practical interest concerns differences between particular characteristics of the population density functions. The $j^{th}$ moment corresponding to the probability density function for the population in recession i is defined by

$$m_{ij} = \int_{-\infty}^{\infty} x^j f_i(x)dx, \qquad (2)$$

where $f_i(x) = dF_i(x)/dx$ is the probability density function of the population of price changes during recession i. When j=1, $m_{i_1} = \mu_i$, the mean price change during recession i. The variance of price changes during recession i is given by $\sigma_i^2 = m_{i_2} - m_{i_1}^2$. Denote by $\underset{\sim}{m}_i$ the vector of moments for recession i, $\underset{\sim}{m}_i = (m_{i_1}, m_{i_2} \ldots)$. Formally, the second question of interest is: Do the price-change data contain any evidence against some null hypothesis of the form

$$H_0': g(\underset{\sim}{m}_{i_1}) = g(\underset{\sim}{m}_{i_2}) = \ldots = g(\underset{\sim}{m}_{i_n}). \tag{3}$$

Here g is a scalar-valued function, $i_1$, $i_2$ ... $i_n$ are unique and $n \leq k$.

Choice of a test procedure for a null hypothesis of form (1) or form (3) should be made only after consideration of the alternative hypothesis of particular interest. The size of a hypothesis-testing procedure, or the type I error probability associated with the procedure, is the probability that the procedure will lead to rejection of a null hypothesis that is in fact true. The power of a test is the probability that the null hypothesis will be rejected when it is actually false. Among a group of test procedures of the same size a desirable procedure is one with high power against the alternative of particular interest. Such a procedure will reject the null hypothesis with high probability if the alternative is true.

Note that thus far nothing has been said about the particular form of the distribution functions for the population of price changes. Parametric tests of hypotheses of form (1) or form (3) involve assumptions that the distributions involved are members of particular parametric classes (normal or chi-squared, for example). Nonparametric test procedures do not require such assumptions. Parametric tests are more powerful than nonparametric procedures provided that the distributional assumptions they involve are correct. When these assumptions are incorrect, however, both the size and the power of a parametric test procedure can differ substantially from their theoretical values. Parametric tests are not particularly robust to departures from the assumptions. In the current context there is no a priori evidence that the density functions of price changes belong to any parametric class. Subsequent attention will focus on nonparametric test procedures.

In the literature (e.g., Durbin, 1973, pp. 39-47) there are a number of tests for hypotheses of form (1) against the alternative that at least two of the cumulative density functions differ in an unspecified way. These tests are not particularly useful in the present context for two reasons. First, it is difficult to provide an economic interpretation of the results of such a test given the vague definition of the alternative hypothesis. The other objection is statistical. To illustrate this objection consider the simple hypothesis

$$H_0: F_1(x) = F_2(x), \text{ for all } x. \tag{4}$$

The set of price changes for the available Consumer Price Index series during recession 1 is a sample of observations from cumulative density $F_1$. The nonparametric tests described by Durbin require a sample of observations from $F_2$ that is independent of the sample of observations

from $F_1$. In our practical context we get a sample of observations from $F_2$ by computing a measure of price change during recession 2 for each of the series used to obtain the sample from $F_1$. One would expect that if a particular price drops relatively far (compared to other prices) during recession 1 it should also drop relatively far during recession 2. Hence our sample of observations from $F_1$ is not independent of the sample from $F_2$. For these reasons, tests of hypotheses of form (1) will not be further considered here.

We will now consider tests of hypotheses of form (3). A simple case of such a hypothesis suggests that the means of the probability density functions for the populations of price changes are equal during two recession periods:

$$H_0': u_i = u_j. \tag{5}$$

Two nonparametric test procedures for (5) against the alternative

$$H_1': u_i \geq u_j \tag{5a}$$

will be discussed. Unlike parametric tests, these procedures do not require any assumptions about the form of the cumulative density function for recession i, $F_i$. However, it is necessary to assume that $F_j$ differs from $F_i$, whatever the form of the latter, only in its mean. This assumption is sufficient for use of one nonparametric test procedure, the sign test. In order to use the second nonparametric procedure, the Wilcoxon signed rank test for paired samples, it is necessary to make the additional assumption that the probability density functions $f_i$ and $f_j$ are symmetric about their means. For the cumulative density function such symmetry implies that

$$F_i(\mu_i - x) + F_i(\mu_i + x) = 1, \text{ for all x.} \tag{6}$$

The Wilcoxon test is more powerful than the sign test if the true distributions have these symmetry properties, but it should not be used in the absence of symmetry.

It is important to note the implications of the assumption required by both the sign test and the Wilcoxon test. Essentially these are tests of a hypothesis of form (1) with respect to a specific alternative; namely, that the difference between the cumulative density functions is due exclusively to a difference between the means of the distributions. We will return later to a discussion of the implications of violation of this assumption.

Computation of a sign test for (5) is straightforward. Suppose that m price series are available. It is necessary to determine the number of

prices that do not drop as quickly in recession i as in recession j. Suppose that there are s such series. The number s should be compared to the distribution of this quantity when the null hypothesis is true, the binomial distribution with m trials and success probability 1/2. Values of s greater than the 100 $\alpha$% point of the binomial cumulative density function provide evidence against (5) in favour of (5a) at the 100 $\alpha$% level of significance. Appropriate tables of the binomial distribution can be found in Pearson and Hartley (1976a, pp. 210-211).

To compute a Wilcoxon paired rank test of (5), it is necessary to calculate, for each price, the difference between its change in recession i and in recession j. These differences must then be ranked from 1 to m, in order of increasing absolute value. The test procedure uses the sum of the ranks associated with positively signed differences. The distribution of this sum when (5) is true is tabulated in Pearson and Hartley (1976b, p. 231) for small m. For larger m the distribution can be approximated using the normal distribution (Pearson and Hartley, 1976b, p. 49). Large values of the test statistic relative to this distribution provide evidence against (5) in favour of (5a).

Before discussing the importance of the assumptions underlying the sign test and the Wilcoxon paired rank test we will consider the problem of testing a hypothesis more general than (5). In particular consider

$$H_0': \mu_{i_1} = \mu_{i_2} = \dots = \mu_{i_n} \qquad (7)$$

where $i_1, i_2 \dots i_n$ are distinct and $n \leq k$. Choice of an appropriate test procedure for (7) depends on the alternative hypothesis of interest. Most relevant in the current context is an alternative suggesting a time ordering,

$$H_1': \mu_{i_1} \leq \mu_{i_2} \leq \dots \leq \mu_{i_n}. \qquad (8)$$

Three test procedures for (7) that are particularly powerful against alternative (8) are discussed, all of which operate in general as follows. For each price series one selects the subset of the data on price changes related to recessions $i_1, i_2 \dots i_n$ and computes a statistic. These statistics are then added across all m prices and compared to the distribution of the sum computed assuming $H_0'$ is true. The distributional assumptions required for the sign test are necessary in each case. The extra symmetry assumption needed to use the Wilcoxon paired rank statistic is not necessary.

The first test procedure uses the Spearman coefficient of rank correlation. For each price series, the changes in the price during the recession periods involved in the hypothesis are ranked from 1 to n in ascending order. From each of these ranks the rank corresponding to the

time-ordering of the relevant recession is subtracted. Spearman's coefficient of rank correlation is then computed.

$$R = 1 - 6 \sum_{i=1}^{n} d_i^2/(n(n^2-1)) \tag{9}$$

where $d_i$, $i=1,2 \ldots n$ are the rank differences. There are m of the statistics, one for each price. The hypothesis (7) can be tested versus alternative (8) by adding the m calculated values of R and comparing the sum to its distribution computed assuming (7) is true. Values of the sum greater than the $100(1-\alpha)\%$ point of this distribution provide evidence against (7) in favour of (8) at the $100(1-\alpha)\%$ level of significance. The distribution of (9) is tabulated (e.g., Gibbons, 1976, pp. 417-418). Writing a computer program to compute the distribution of a sum of m independent Spearman rank correlations, each based on n observations, is a simple computational task.

The other two test procedures for (8) are analogous. One procedure uses Kendall's tau ( $\tau$ ), a rank correlation coefficient. It is necessary to rank the set of relevant changes in each price in ascending order. Then this set of ranks should be arranged according to the time-ordering of the recessions. For example suppose that there are three recessions; the most negative price movement occurs in recession 2, the most positive movement in recession 3. The appropriate arrangement of ranks is $\{2,1,3\}$. To compute the Kendall's tau it is necessary to consider every ordered pair of ranks drawn from this set, namely (2,1), (2,3) and (1,3). For each ordered pair the second rank is subtracted from the first. Suppose that there are u positive differences. (In the example u = 2.) The Kendall coefficient of rank correlation is

$$\tau = 1-4u/(n(n-1)). \tag{10}$$

The distribution of (10) when (7) is true is tabulated in Gibbons (p. 420). To test (7) it is necessary to add the tau statistics computed for each price and compare the sum against the distribution of a sum of m independent Kendall tau statistics, each based on n observations, computed assuming (7) is true. Unusually large values of this statistic provide evidence against (7) in favour of (8). The distribution of a sum of m independent Kendall tau statistics is not tabulated but can be computed with relative ease.

The third test procedure is based on the peak test originally proposed by Goldfeld and Quandt (1965) to test for the heteroscedasticity of regression residuals. For each price, the relevant price-change measures are arranged according to the time-ordering of the recessions. This sequence is denoted by $\{p_1, p_2 \ldots p_n\}$. In this sequence

$p_j$ is a peak if $j \neq 1$ and, for all $i<j$, $p_i<p_j$. It is necessary to count the number of peaks that occur for each price. These statistics are added, and the sum is compared to the appropriate distribution computed assuming that the null hypothesis is true. Goldfeld and Quandt tabulate density functions for a single peak statistic when (7) is true. Some computer programming is necessary to calculate the distribution of a sum of such statistics. Unusually high values of this sum provide evidence against (7) in favour of (8).

Alternative hypothesis (8) is called a composite hypothesis because there are a large number of cumulative density functions, $F_{i_1}, F_{i_2}, \ldots F_{i_n}$, for which it is satisfied. When cumulative density functions are specified, except for some unknown parameters, for most tests of a simple hypothesis against a composite alternative there is one test procedure of a certain size that is more powerful than all others of the same size. This is not the case in a nonparametric context. Each of the three test procedures described above is most sensitive to slightly different choices of $F_i$'s satisfying (8). Use of more than one test statistic for (7) should increase the chances of rejecting the null hypothesis, $H_0$, when it is in fact incorrect.

Recall the assumption about the population distributions of price changes introduced at the beginning of this discussion. To test for differences between the means of distributions it is necessary to assume that, regardless of their parametric form, the distributions differ only in their means. What are the implications of violation of this assumption? A test procedure may lose power. The most important problem is that it is no longer possible to control the type I error of a test procedure. If actual distributions have the same means but differ in terms of variance, skewness or other characteristics, true null hypotheses concerning means may be rejected more than $100\alpha\%$ of the time using a test of size $\alpha$.

What should be done to guard against the undesirable effects of violation of the assumption? Hawkins (1980) has a suggestion. Before using nonparametric tests for means, one should adjust each sample of data -- in this case each group of observed price changes from a particular recession -- so that all samples have the same sample variance. This can be achieved by dividing each sample through by a constant determined using the relationship $Var(ax) = a^2Var(x)$. In principle it would be possible to perform further adjustments to correct for differences in sample skewness and kurtosis. Note that these corrections are only approximate since they equate <u>sample</u> characteristics rather than population characteristics. The properties of the corrections have not been much investigated in the statistical literature. Lehmann (1975) demonstrates that use of a mean correction when testing for variance differences is acceptable in the sense that type I error probabilities can be controlled, if the samples involved are large enough. Apparently no one has examined

the use of a variance correction before testing for equality of means. In practice, test statistics should be computed with and without this correction in order to obtain some indication of the importance of the assumption for the test results.

Finally it is necessary to address the issue of testing for hypotheses concerning characteristics of population distributions of price changes other than their means. Such tests can be conducted using any of the test statistics described above after an appropriate initial transformation of the data. To test for equality of variances each price-change measure should be squared before the procedures described above are applied. Skewness and kurtosis can be examined using third and fourth powers. Remember that in each case one assumes that the price-change distributions under study are identical except for the characteristic of interest. Sample mean and variance corrections should be used whenever relevant to check the importance of this assumption for test results.

## REFERENCES


Armstrong, J. "Testing for Changes in the Cyclical Sensitivity of Prices: Methodological Comments." Research Department, Bank of Canada, RN-83-253, 1983.

Cagan, P. "Changes in the Recession Behaviour of Wholesale Prices in the 1920s and Post-World War II." Explorations in Economic Research, 2, 1975, pp. 54-104.

Durbin, J. "Distribution Theory for Tests Based on the Sample Distribution Function. Philadelphia: Society for Industrial and Applied Mathematics, 1973.

Gibbons, J.D. Nonparametric Methods for Quantitative Analysis. New York: Holt, Rinehart and Winston, 1976.

Goldfeld, S.M. and Quandt, R.E. "Some Tests for Heteroscedasticity." Journal of the American Statistical Association, 60, 1965, pp. 539-547.

Hawkins, D.M. Identification of Outliers. London: Chapman and Hall, 1980.

Lehmann, E.L. Nonparametrics: Statistical Methods Based on Ranks. San Francisco: McGraw-Hill, 1975.

Pearson, E.S. and Hartley, H.O. (eds.). Biometrika Tables for Statisticians, volume 1, Cambridge: Cambridge University Press, 1976a.

Pearson, E.S. and Hartley, H.O. (eds.). Biometrika Tables for Statisticians, volume 2, Cambridge: Cambridge University Press, 1976b.